Catalytic crosslinking-based methods for enzyme-specified profiling of RNA ribonucleotide modifications

Shobbir Hussain*

Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

Keywords: *epitranscriptome; epitranscriptomics; RNA methyltransferase; Aza-IP; miCLIP; m5C; m2A; m5U; pseudouridine*

Correspondence: s.hussain@bath.ac.uk

Abstract

Well over a hundred types of naturally occurring covalent modifications can be made to ribonucleotides in RNA molecules. Moreover, several types of such modifications are each known to be catalysed by multiple enzymes which largely appear to modify distinct sites within the cellular RNA. In order to aid functional investigations of such multi-enzyme RNA modification types in particular, it is important to determine which enzyme is responsible for catalysing modification at each site. Two methods, Aza-IP and methylation-iCLIP, were developed and used to map genome-wide locations of methyl-5-cytosine (m5C) RNA modifications inherently in an enzyme specific context. Though the methods are quite distinct, both rely on capturing catalytic intermediates of RNA m5C methyltransferases in a state where the cytosine undergoing methylation is covalently crosslinked to the enzyme. More recently the fundamental methylation-iCLIP principle has also been applied to map methyl-2-adenosine sites catalysed by the *E.Coli* RlmN methylsynthase. Here I describe the ideas on which the two basic methods hinge, and summarise what has been achieved by them thus far. I also discuss whether and how such principles may be further exploited for profiling of other RNA modification types, such as methyl-5-uridine and pseudouridine.

Introduction

The canonical nucleotides in RNA molecules can undergo various types of stable biochemical modification, including but not limited to methylation (Machnicka et al. 2013; Boccaletto et al. 2018). The majority of known modification types have been identified in all three kingdoms of life and are thought to be of significant biological importance (Towns and Begley 2012). Of the 100+ types of RNA modifications, some of their genomic co-ordinates have been determined precisely though not always comprehensively; nonetheless a core view of their distribution is already apparent (Grosjean 2015). Transfer RNAs (tRNAs) are the most heavily modified RNA biotype with different types of modifications occurring every few nucleotides; stability, amino-acylation, and translation fidelity are just a few of the crucial functions that are directly attributable to some individual tRNA modifications (El Yacoubi et al. 2012). Modifications on ribosomal RNAs (rRNAs) which influence their stability (Sharma and La Fontaine 2015), as well as modifications occurring on messenger RNAs (mRNAs) which can influence turnover or rates of translation (Zhao et al. 2017), are also some notable examples of key functional roles already demonstrated.

Much of our knowledge relating to the enzymatic factors that catalyse some of these RNA ribonucleotide modifications originates from studies performed in prokaryotes, archaea and single-celled eukaryotes (Grosjean and Benne 1998). A common theme that clearly emerged even from such early investigations is that a single modification type is very often catalysed by multiple enzymes, and further that such enzymes will only act on specific sites in the cellular RNA (Ferre-D'Amare 2003). Perhaps two of the most notable examples of such modification types are methyl-5-cytosine (m5C) and pseudouridine. Indeed in mammals, m5C in cellular RNA is catalysed by at least eight enzymes (NSun1-7, Dnmt2), and the target RNA sites for a few of these have been surveyed in a genome-wide context. The previously incorrectly suspected DNA methyltransferase, Dnmt2, has been found to methylate cytosines at position 38 of a few tRNA isoacceptors (Schaefer et al. 2010; Khoddami and Cairns 2013). NSun2 on the other hand methylates cytosines predominantly located at positions 48-50 in tRNAs (Khoddami and Cairns 2013; Hussain et al. 2013a), and also appears to be capable of methylating a few sites in other ncRNAs as well as potentially in mRNAs (Squires et al. 2012; Khoddami and Cairns 2013; Hussain et al. 2013a; Hussain et al. 2013b). NSun3 catalyses methylation at position C34 of mitochondrial tRNAs (van Haute et

al. 2016) whereas NSun6 methylates C72 of a few cytoplasmic tRNA isoacceptors (Haag et al. 2016). Using an *NSun4* knockout transgenic mouse control coupled with a bisulfite sequencing approach, it was also demonstrated that the NSun4 enzyme catalyses methylation of C911 of mouse 12S rRNA (Metodiev et al. 2014). The target sites of a few of the 13 known mammalian pseudouridine synthases have been profiled, and such sites are also found to be just as enzyme-specific (Li et al. 2015; Safra et al. 2017; Zaganelli et al. 2017).

Whereas the enzyme specificities of some pseudouridine sites have so far been determined by performing genetic perturbations, those for RNA m5C sites have instead been established largely using substrate-trapping techniques. The two current methods which enabled this are 5-azacytidine-mediated IP (Aza-IP) (Khoddami and Cairns 2013) and methylation-iCLIP (miCLIP) (Hussain et al. 2013a). Though distinct, both techniques work by stalling the normal catalytic process at a step where a key residue in the enzymatic active site becomes covalently crosslinked to the cytosine undergoing modification. The enzyme-substrate catalytic intermediate complexes are purified, and the precise crosslink/modification sites then determined using next generation sequencing-computational approaches.

Further to helping understand how the Aza-IP and miCLIP techniques work mechanistically, an appreciation of the relevant catalytic mechanism will also help establish the potential scope of the method principle i.e. its suitable application to other modification types. Here I thus begin with a brief historical perspective of our knowledge pertaining to enzyme-RNA catalytic adducts. I then describe how the Aza-IP and miCLIP techniques allow nucleotide-resolution determination of RNA m5C sites in a genome-wide context, and what they have helped uncover so far. Finally I propose in more detail how the method principle may be applied to other RNA modification types and particular enzymes.

Capturing catalytic crosslinks in ribonucleotide modification reactions

Several early studies had implicated the formation of a covalently linked complex between various types of pyrimidine modifying enzymes and either target pyrimidines existing in monomeric form or within RNAs (Yeh et al. 1967; Schoemaker and Schimmel 1977a; Schoemaker and Schimmel 1977b; Kunitani and Santi 1980; Santi and Hardy 1987). Such

complexes, referred to by biochemists as a type of Michael adduct (Box 1), were established as representing catalytic intermediates which in a later step of the pathway are resolved to release the fully modified pyrimidine and regenerate the active enzyme. An important example were aminoacylation reactions catalysed by tRNA synthetases, which were shown to proceed via a Michael adduct forming on a universally conserved uridine at position 8 of tRNAs (Schoemaker and Schimmel 1977a; Schoemaker and Schimmel 1977b; Starzyck et al. 1982). It was further demonstrated that a key cysteine residue in the synthetase enzyme likely performs a nucleophilic attack on the 6-carbon of the uracil heterocyclic ring to form the covalent crosslink (Starzyck et al. 1982).

Nucleophilic attack mediated by a cysteine residue from within the enzymatic active site indeed appears to be an established common feature of Michael adduct formation in RNA ribonucleotide modification reactions (Kealey and Santi 1991), including those mediated by RNA m5C methyltransferase enzymes (Liu and Santi 2000). According to the established enzymatic scheme, covalent bond formation between the key cysteine residue in RNA m5C methyltransferases and the 6-carbon of the cytosine heterocyclic ring renders the neighbouring 5-carbon susceptible to electrophilic attack. A methyl group can then be readily transferred to the 5-carbon before abstraction of its hydrogen. Elimination of the enzyme from the 6-carbon is the final step of the scheme (Figure 1A). It was proposed (Santi et al. 1983), and later demonstrated (Wu and Santi 1985), that presence of a nitrogen atom in place of the carbon at position 5 of the base ring i.e. as occurs in 5-azacytosine, would stall the enzymatic process. In such a scenario, although normal enzyme-substrate crosslinking occurs, the subsequent step of methyl group addition to position 5 cannot occur thus yielding a stabilised Michael adduct (Figure 1B). Presence of 5-azacytosine in place of cytosine in RNA would thus represent a potential mechanism for capturing crosslinked m5C methyltransferase-cytosine substrate complexes. The Aza-IP method is centred on this principle.

An unexpected finding further made during the aforementioned investigations of the relevant enzymatic mechanism, was that although DNA m5C methyltransferases and RNA m5C methyltransferases share common catalytic motifs, they are observed to employ distinct conserved cysteine residues as the nucleophilic catalyst (Liu and Santi 2000). Particularly puzzling was the fact that the pivotal cysteine residue used by DNA m5C

methyltransferases as the nucleophile was found to be just as highly conserved in RNA methyltransferases but apparently not required for methyl group addition to RNA substrates. It was later shown that this second cysteine residue in RNA methyltransferases is indeed also involved in the catalytic process, but in fact at a step following the methyl group addition (King and Redman 2002; Redman 2006), making it likely that it acts as the general base required for abstraction of the hydrogen from the methylated 5-carbon (Lee et al. 2005). Thus substitution of this cysteine could prevent hydrogen abstraction and also the subsequent elimination of the enzyme from the methylated substrate (Figure 1C), thereby offering a means of capturing the Michael adducts in these reactions. The miCLIP method is based on such principles.

Aza-IP and miCLIP for genome wide profiling of RNA ribonucleotide modifications

5-azacytidine mediated RNA immunoprecipitation (Aza-IP)

The Aza-IP technique was developed and initially used for genome wide profiling of RNA m5C modifications catalysed by human NSun2 and Dnmt2 (Khoddami and Cairns 2013). Full protocol details and relevant technical commentaries have been published elsewhere (Khoddami and Cairns 2014). Briefly, the methodology involved culturing cells in the presence of 5-azacytosine in order to enable incorporation of the nucleotide analogue into RNAs during their transcription. Ectopic expression of epitope-tagged Dnmt2 or NSun2 enzyme was performed, and antibodies complementary to the epitope tag were then used to purify the enzyme-RNA Michael adducts. Early in the library preparation, these complexes then dissociate in solution resulting in a 5-azacytosine ring in an open conformation which will base pair not with guanosine but instead with cytosine (Jackson-Grusby et al. 1997); thus in synthesized cDNAs, G to C substitutions are expected at enzyme-targeted sites. PCR-amplified libraries are prepared and then run on a high-throughput Illumina sequencer. Finally, computational algorithms are used to determine positions of G to C substitutions compared to reference sequence, thus enabling RNA m5C modification sites to be inferred precisely at nucleotide-resolution.

Cytosine 38 of a few tRNA isoacceptors were previously characterised as target sites for the Dnmt2 m5C methyltransferase (Schaefer et al. 2010), and such findings were recapitulated

in the Aza-IP study. The previously known targets of NSun2 was limited to only a few tRNA isoacceptors, with positions 48, 49, and 50 most frequently identified as the target sites. The Aza-IP study not only confirmed all previously known such sites, but also showed that several more tRNA isoacceptors were in fact methylated at the 48-50 positions. The enrichment-based genome-wide approach had therefore succeeded in yielding a significantly more comprehensive map of NSun2-target m5C sites in tRNAs. The study also unveiled a few novel m5C sites located in other RNA biotypes including in 5S rRNA, Vault RNAs, SCARNA2, Y RNA and 7SL RNA. Although several potential NSun2-depedent target sites in mRNAs were also identified by Aza-IP, only one of these (located in the SHF mRNA) passed the rigorous statistical criteria that was applied to accurately call targeted tRNA sites too.

Methylation-individual nucleotide resolution crosslinking immunoprecipitation (miCLIP)

The miCLIP method was developed and also initially used to map the genome wide locations of NSun2 in RNAs (Hussain et al. 2013a). Full protocol details along with relevant notes and discussion have been published elsewhere (George et al. 2017). Briefly, the approach involved ectopic expression of an epitope-tagged point-mutant form of NSun2. Here the catalytic cysteine thought to act as the general base, and thus critical for allowing release of the enzyme from the methylated substrate, was mutated to alanine. The stabilised Michael adducts were then purified using an antibody complementary to the epitope tag. In this methodology, the crosslink which forms on natural target cytosines are more permanently stable, and can be subjected to denaturing polyacrylamide gel electrophoresis thus enabling purification of enzyme-substrate complexes at exceptionally high stringency. Reverse transcriptions of isolated substrate RNAs stall at the Michael adduct crosslink sites, and thus the final nucleotide of the synthesized cDNAs are predicted to correspond to the methylated cytosine position. The use of unique molecular identifiers (UMIs) within the primers used for RT during library preparation allow for PCR duplicate reads to be discarded, thus enabling mapping of the high-throughput sequence reads at the cDNA-level.

The NSun2 miCLIP study mirrored the major findings of Aza-IP, in that opposed to just a small handful, most human tRNA isoacceptors were in fact found m5C modified at the

48/49/50 positions. Further, the novel NSun2-dependent m5C sites in the Vault ncRNAs and 5S rRNA reported by the Aza-IP study were also detected by miCLIP. The miCLIP study also reported several NSun2-dependent mRNA target sites (Hussain et al. 2013a; Hussain et al. 2013b), including the SHF mRNA site which reached the statistical cut-off applied in the Aza-IP study, though these remain to be cross-validated using additional methods. Some of the NSun2 miCLIP-detected sites in ncRNAs were however additionally validated and further investigated in the study. For example, mutations in the human *NSun2* gene had previously been established as causative in Mendelian intellectual disability syndromes in some families (Abbasi-Moheb et al. 2012; Khan et al. 2012); bisulfite sequencing of RNA derived from dermal fibroblasts of unaffected siblings were used to confirm the presence of miCLIP-detected m5C sites in Vault RNAs, whereas these methylations were found to be absent in bisulfite-sequenced Vault RNA derived from affected patient cells (Hussain et al. 2013a).

Subsequent further successful applications of Aza-IP and miCLIP

Aza-IP has more recently been applied to determine the genome-wide profile of NSun6 m5C target sites (Haag et al. 2015), and it was demonstrated that NSun6 targets position 72 of cysteine and threonine tRNA isoacceptors in a manner dependent on the presence of their CCA tails. miCLIP has further been applied to determine the targets of the NSun3 RNA methyltransferase, and it was found that the enzyme targets position 34 of mitochondrial methionine tRNA (Van Haute et al 2016). The study also demonstrated that NSun3 mutations are causative in an autosomal recessive form of mitochondrial disease characterised by developmental disability and oxidative phosphorylation deficiency in skeletal muscle. Mitochondria of patient-derived dermal fibroblasts were found to have markedly decreased levels of protein synthesis rates compared to that within mitochondria from matched control cells. Since the miCLIP-detected m5C site at C34 of mitochondrial methionine tRNA was previously reported to be a site carrying the 5-formylcytosine (f5C) modification, it was further hypothesized that m5C at this site might represent an intermediate which is further metabolised to form a final f5C at this position. Consistent with this idea, it was indeed found that the NSun3-defecient dermal fibroblasts lacked f5C at the mitochondrial methionine tRNA C34 site, whereas it was readily detected at this position in matched controls. In summary, the NSun3 study identified the presence of m5C

at the C34 'wobble' position of mitochondrial translation initiator tRNA, which was shown to be necessary for formation of f5C at this position, and likely thus for efficient mitochondrial protein synthesis. Disruption of the NSun3 enzyme, which catalyses the initial step of the modification scheme, results in a characteristic mitochondrial disease phenotype in humans.

Much more recently, the methylation-iCLIP catalytic crosslinking principle was applied to map methyl-2-adenosine (m2A) modifications catalysed by the *E.Coli* RNA methylsynthase, RlmN (Stojkovic et al. 2018). The method termed miCLIP-MaPseq mainly differs from the original miCLIP technique in that it employs the thermostable group II reverse transcriptase (TGIRT) for synthesis of cDNAs during library preparation. As TGIRT has a high tendency to read through crosslink sites, cDNA truncations are not observed, but instead mutations specifically at the crosslink site are frequently inserted in the cDNAs during synthesis. This then enables precise target sites in RNAs to be identified via mutation screening within the mapped cDNA libraries. The study successfully mapped RlmN target sites in 23S rRNA as well as showing that the enzyme can methylate several tRNA isoacceptors.

Shortcomings and scope for improvement of the two techniques

Though both miCLIP and Aza-IP have unveiled several novel RNA m5C sites, some of which have further been demonstrated to be have significant biological significance, both techniques suffer from shortcomings.

For Aza-IP, perhaps the main one, is that 5-azacytosine is also incorporated into DNA during replication and inhibits DNA methylation via covalent trapping of DNA methyltransferases. When the aim of an investigation becomes determining the biological roles of m5C RNA modifications, due consideration must therefore be given to the impact that the unwanted DNA methylation inhibition might have on the biological phenomena being studied. This may be particularly important when studying biological processes known to hinge on epigenetic changes such as early developmental or cancer disease states for example. A much more minor shortcoming of the Aza-IP approach is that it may not be relatively effective in detecting low-abundance modifications. This is since 5-azacytosine will compete with cytosine for incorporation into RNA during transcription: thus for each reference cytosine site, some transcripts will contain 5-azacytosine, but others which might be some

significant proportion, will contain cytosine. Thus a methylation attempt would need to be made specifically on the transcripts containing 5-azacytosine at the correct position for crosslinking to occur. For RNAs methylated at high stoichiometry this is not an issue, but low abundance modifications, such as those that might occur on mRNAs for example, may fall below the limits of detection.

For miCLIP, the main disadvantage is the need to ectopically express the relevant cysteine point mutant form of the enzyme, which could potentially lead to the detection of non-physiological off-targets. The ideal m5C-profiling catalytic crosslinking experiment, perhaps, would be a miCLIP approach where the relevant cysteine mutation is knocked into the RNA methyltransferase gene via editing. As alluded to earlier, catalytic crosslinking has the likely potential to be used to profile some various types of RNA ribonucleotide modifications; an Aza-IP-type approach applied to detect uridine modifications (discussed in further detail in the following section) would also be particularly attractive, as there is no longer the concern of DNA methylation inhibition.

In summary, though the Aza-IP and miCLIP techniques can detect RNA modification sites with a very good degree of confidence, when used in their current form, a few important caveats as summarised above should be considered.

Applying catalytic crosslinking principles to profile other RNA modifications

Considering especially the recent successful application of the fundamental methylation-iCLIP catalytic crosslinking principle to map an adenosine modification (Stojkovic et al. 2018), it is conceivable that substrate trapping could indeed be employed to map a variety of other RNA modification types. As discussed previously, key to nucleotide modification catalytic crosslinking is the existence of a central amino acid residue which forms a covalent bond with the target substrate; in some cases a second key catalytic residue is critical for resolution of the covalent complex, but in other cases, such as for DNA methyltransferases for example, no such second critical residue can be distinguished. Thus when deciding whether Aza-IP- or miCLIP-type experiments might be suitable for mapping a particular RNA modification type, a couple of broad guidelines may be posited: i) modifications identified to involve two key residues for catalysis, one for covalent bond formation and the other for

resolving the covalent intermediate, may be amenable to both an Aza-IP- and miCLIP-type approach; ii) modifications thought to involve one key catalytic residue i.e. for covalent bond formation, may be amenable only to a potential Aza-IP-type methodology. To provide a more detailed illustration, below I work through a couple of examples of uridine modifications which may be amenable to catalytic crosslinking-based mapping.

Methyl-5-uridine (m5U)

A good candidate for catalytic crosslinking-based profiling is perhaps methyl-5-uridine (m5U): like m5C, a pyrimidine modification which involves methylation at the 5-carbon position of the base ring. In support of this, it has previously been shown that bacterial tRNA m5U methyltransferase proceeds via a catalytic mechanism similar to that described for m5C methyltransferases (Santi and Hardy et al. 1987; Kealey et al. 1994), thus being dependent on a catalytic cysteine nucleophile for Michael adduct formation (Kealey and Santi 1991). During such studies it was found that presence of the uridine analogue 5-flourouracil (5FU), where the hydrogen at the 5-carbon is substituted with a fluorine atom, in RNAs results in covalent trapping of m5U methyltransferases. In such cases, the catalytic process proceeds to include methylation of the 5-carbon, however because of the stability of the carbon-fluorine bond, the enzymatic scheme stalls yielding a stabilised Michael adduct (Figure 2). The use of 5FU, which is readily incorporated into cellular RNA during transcription, may thus enable m5U profiling i.e. akin to the use of 5-azacytosine to profile m5C positions in cellular RNA.

It has also been shown that the m5U methyltransferases behave like most RNA m5C methyltransferases, thus requiring a second key catalytic residue to resolve the catalytic adducts and complete the enzymatic process (Lee et al. 2005; Alian et al. 2008). However, unlike the case with RNA m5C methyltransferases, the second central catalytic residue in m5U methyltransferases was in fact found to be a glutamate, not a cysteine. It was shown that the catalytic glutamate likely acts as the general base required for abstraction of the hydrogen from the already methylated 5-carbon (Lee et al. 2005). Thus, an inhibition of this penultimate step would in turn result in prevention of the final step, i.e. elimination of the enzyme from 6-carbon of the substrate uridine. It follows that mutation of this conserved

catalytic glutamate residue may potentially enable a relevant miCLIP-type experiment to be performed.

Bacteria and yeast harbour a single known m5U methyltransferase enzyme, TrmA and Trm2 respectively. These are known to universally target position U54 of the T-loop in tRNAs. In mammals, two m5U methyltransferases TRMT2A and TRMT2B, are predicted based on sequence homology. Experimentally determined target sites for either enzyme are however yet to be reported. It is noteworthy that, apart from a few of the key residues, there is only very moderate sequence homology within the catalytic domains of the two enzymes; they likely thus potentially target distinct sites in the cellular RNA. Aza-IP-type and/or miCLIP-based experiments may be informative here.

Pseudouridine

The isomerisation of uridine to pseudouridine is known to be catalysed by families of pseudouridine synathases in the three kingdoms of life. These can be grouped according to phylogenetic association with the five bacterial enzymes, TruA, TruB, TruD, RluA and RsuA, which modify distinct sites mostly found in tRNAs, rRNAs and snRNAs. Similar to the observations with m5U methyltransferases, some bacterial pseudouridine synthases have also been found to form stable complexes when incubated with tRNA substrates containing 5FU in place of uridine ribonucleotides (Huang et al. 1998; Gu et al. 1999). However, unlike RNA m5C and m5U methyltransferases, pseudouridine synthases utilise an aspartate residue, as opposed to a cysteine residue, to form a covalent attachment to the uridine. The precise chemical process via which the covalent attachment occurs and its exact place in the enzymatic scheme, is however as of yet, not fully clear (Spedaliere et al. 2004). To add to the uncertainty, details of the catalytic mechanism employed may potentially differ for pseudouridine synthases belonging to different phylogenetic families (Hamma and Ferre-D'Amare 2006). Nonetheless, there is some evidence which suggests that at least two distinct bacterial pseudouridine synthases, TruA and RluA, do most likely catalyse their reactions via covalent attachment to the uridine substrate and that the crosslinks can indeed be stabilised via substitution with 5FU in targets (Huang et al. 1998; Gu et al. 1999; Spedaliere and Mueller 2004).

Of the 13 known mammalian pseudouridine synthases, the target sites of only a few have been profiled in a genome or transcriptome-wide context. These include the TruA family member PUS1, TruB members TRUB1 and TRUB2, and the TruD family member PUS7 (Li et al. 2015; Safra et al. 2017). These studies, which relied on genetic perturbation of the respective enzymes via gene disruption or siRNA-mediated knockdown, indicated that TRUB1, PUS7 and PUS1 can all catalyse uridine isomerisation at specific sites in human mRNAs. Additionally, two novel PUS1-dependent pseudouridine sites in the SRA lncRNA were also identified. A protein-RNA interaction based genome-wide screen for the RsuA/RluA member RPUSD4 has also been performed via the HITS-CLIP strategy, and aided in revealing that the enzyme targets mitochondrial tRNAs (Zaganelli et al. 2017). Collectively, such findings indicate that all mammalian pseudouridine synthases might act on distinct sites within the cellular RNA as opposed to having redundant enzymatic roles. Thus enzyme-specific genome-wide profiling of pseudouridine modification in mammalian systems will likely help uncover new important insights into its wider functional roles. Based on current observations, the RluA/RsuA members RPUSD1-4 and the TruA members PUS1, PUSL1, and PUS3 may be considered prime candidates for attempts at an Aza-IP-type pseudouridine profiling method, where 5FU is employed for catalytic crosslinking stabilisation.

Concluding remarks

Though enzyme-specified profiling of RNA ribonucleotide modifications can currently involve genetic perturbation, to perform such experiments reliably requires gene disruption which is labour intensive. Such an approach also relies on comparing two conditions i.e. wt vs ko, meaning that identified sites are much more susceptible to the details of the computational analysis workflow/statistical models employed to call differential expression. Though Aza-IP- and miCLIP-type profiling currently have their shortcomings too, the ability to profile specific enzymatic targets inherently may be considered overall a distinct advantage. Given the current pervasive observations of unique site-specificity of RNA modification enzymes, further improvement and developments of the two techniques will

likely render them as ideally placed to more informatively profile some types of RNA ribonucleotide modification in future.

Acknowledgements

Research in the author's laboratory is funded by the Biotechnology and BioSciences Research Council UK (BBSRC).

Box 1. ***The Michael addition reaction and Michael adducts***

Initially described by Arthur Michael in 1887, in the simplest of terms, the Michael reaction can be viewed as a widely occurring atom-economical method of carbon-carbon bond formation (Michael 1887). More recent definitions heed recognition that common to any Michael reaction is the addition of a nucleophile to an unsaturated carbonyl compound. Accordingly, the mechanism is also found to be widely adopted in natural reactions for the formation of asymmetrical carbon bonds, including carbon-nitrogen, carbon-oxygen, and carbon-sulphur. Thus although the nucleophile is commonly a carbanion, it may also be, for example, the sulphur-containing group of a cysteine amino acid residue. In some enzymatic reactions, the nucleophilic sulfhydryl group of a cysteine residue within the enzymatic active site can therefore participate directly in a Michael reaction, and form a bond via its sulphur with a suitable carbon in the target. As an example, within the heterocyclic ring of pyrimidine base groups, the 6-carbon is particularly susceptible to nucleophilic addition, and thus in some pyrimidine modification reactions, cysteine sulphur:base ring 6-carbon Michael addition reactions can be observed. In such examples however, the carbon-sulphur bond is only an intermediate in the full enzymatic reaction, and is thus eventually broken. The intermediate conjugate in these and similar cases has often been referred to as a 'Michael adduct' (Starzyk et al. 1982; Santi and Hardy 1987).

Figure legends

Figure 1. ***Enzyme-substrate covalent intermediates in RNA m5C modification reactions, and their capture.*** (A) The heterocyclic ring of the base group of a cystosine ribonucleotide is shown. For the sake of clarity, less relevant details of the chemical structure (positions of nitrogen atoms within the ring etc.) have been omitted. RNA m5C methyltransferase enzymes are known to harbour a catalytic cysteine residue capable of mediating a nucleophilic attack (:Enz) of the carbon at the 6 position of the ring. This enables a covalent bond to form between the catalytic cysteine of the enzyme and the 6-carbon, which in turn renders the 5-carbon of the ring susceptible to electrophilic attack. A one carbon group, $^{+}CH3$, (provided by an S-Adenosyl Methionine donor; not shown here) can then form a covalent linkage with the 5-carbon. The proton from the 5-carbon is then abstracted in likely an enzyme dependent manner, and finally, β-elimination from the methylated ribonucleotide regenerates the free active enzyme. (B) In 5-azacytosine, position 5 of the ring harbours a nitrogen instead of the usual carbon. Although covalent linkage of the enzyme to the 6-carbon occurs as normal, the methyl group does not readily attack the 5-nitrogen. The catalytic reaction thus stalls prior to the methylation. (C) In the normal reaction, a second catalytic cysteine residue likely acts as the general base required for proton abstraction from the 5-carbon following its methylation. A mutated form of the enzyme where this key cysteine has been substituted, :Enz-CBM (enzyme-catalytic base mutant), is still able to form the normal covalent linkage to the cytosine. However, the enzymatic scheme stalls at the proton abstraction step, meaning that β-elimination of the enzyme from the methylated ribonucleotide does not occur either.

Figure 2. ***The use of 5-flurouracil to capture enzyme-substrate covalent intermediates in RNA m5U modification reactions.*** In 5-flurouracil, there is a fluorine in place of the usual hydrogen attached to the 5-carbon of the heterocyclic ring of the base. An m5U methyltransferase successfully forms a covalent bond with the 6-carbon during the catalytic reaction. This renders the 5-carbon susceptible to electrophilic attack, and a methyl group is transferred to it. However, due to the particularly stable nature of the 5-carbon:fluorine bond, fluorine abstraction does not occur. The enzymatic scheme therefore stalls, and the m5U methyltransferase remains covalently bound to its substrate ribonucleotide.

References

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., Wieczorek, D., Motazacker, M.M., Esmaeeli-Nieh, S., Cremer, K., et al. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. Am. J. Hum. Genet. 90, 847–855.

Alian et al. 2008. Structure of a TrmA-RNA complex: A consensus RNA fold contributes to substrate selectivity and catalysis in m5U methyltransferases. Proc Natl Acad Sci USA., 105, 6876-81.

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., et al. (2018). MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* 46, D303–D307. doi: 10.1093/nar/gkx1030

El Yacoubi B., Bailly M., de Crecy-Lagard V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu. Rev. Genet, 46:69-95.

Ferre-D'Amare (2003). RNA-modifying enzymes. Curr Opin Struct Biol. 2003 Feb;13(1):49-55.

George H et al. (2017). Illustrating the epitranscriptome at nucleotide resolution using methylation-iCLIP (miCLIP). A. Lusser (Ed.), RNA Methylation, Humana Press (2017), pp. 91-106

Grosjean H. 2015. RNA modification: the Golden Period 1995–2015. RNA **21**:625–626.

Grosjean H, Benne R (Eds): Modification and Editing of RNA. Washington DC: American Society for Microbiology Press; 1998.

Gu, X., Liu, Y., and Santi, D.V. (1999). The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil- tRNA. Proc. Natl. Acad. Sci. USA 96, 14270–14275.

Haag, S. *et al.* NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA* **21**, 1532–1543 (2015).

Hamma T and Ferre-D Amare AR. 2006. Pseudouridine synthases. Chemistry & Biology 13, 1125–1135.

Huang, L., Pookanjanatavip, M., Gu, X., and Santi, D.V. (1998). A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry 37, 344–351.

Hussain, S. et al. (2013a). NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs. Cell Rep. 4, 255–261.

Hussain, S., Aleksic, J., Blanco, S., Dietmann, S., and Frye, M. (2013b). Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. 14, 215.

Jackson-Grusby, L., Laird, P.W., Magge, S.N., Moeller, B.J. & Jaenisch, R. Mutagenicity of 5-aza-2'-deoxycytidine is mediated by the mammalian DNA methyltransferase. Proc. Natl. Acad. Sci. USA 94, 4681–4685 (1997).

Kealey JT and Santi DV. (1991) Identification of the Catalytic Nucleophile of tRNA (m5U54) Methyltransferase. Biochemistry, 30, 9724-9728

Kealey JT, Gu X, Santi DV. Enzymatic mechanism of tRNA (m5U54)methyltransferase. Biochimie. 1994;76(12):1133-42.

Khan, M.A., Rafiq, M.A., Noor, A., Hussain, S., Flores, J.V., Rupp, V., Vincent, A.K., Malli, R., Ali, G., Khan, F.S., et al. (2012). Mutation in NSUN2, which encodes an RNA methyltransferase, causes autosomal-recessive intellectual disability. Am. J. Hum. Genet. 90, 856–863.

Khoddami, V. & Cairns, B. R. (2013) Identification of direct targets and modified bases of RNA cytosine methyltransferases. Nat. Biotechnol. 31, 458–464.

Khoddami, V. & Cairns, B. R. (2014) Transcriptome-wide target profiling of RNA cytosine methyltransferases using the mechanism-based enrichment procedure Aza-IP. *Nat. Protoc.* **9**, 337–361.

King, M.Y., and Redman, K.L. (2002). RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. Biochemistry 41, 11218–11225.

Kunitani, M. G. & Santi, D. V. Biochemistry 19, 1271–1275 (1980).

Lee, T. T., Agarwalla, S., and Stroud, R. M. (2005) A unique RNA fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function Cell 120, 599– 611

Li, X. Y. et al. Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. Nat. Chem. Biol. 11, 592–597 (2015).

Liu Y, Santi DV. (2000) m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. Proc Natl Acad Sci U S A. 97(15):8263-5.

Machnicka, M.A. *et al*. MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Res.* 41, D262–D267 (2013).

Metodiev MD, Spåhr H, Loguercio Polosa P, Meharg C, Becker C, Altmueller J, et al. (2014) NSUN4 Is a Dual Function Mitochondrial Protein Required for Both Methylation of 12S rRNA and Coordination of Mitoribosomal Assembly. PLoS Genet 10(2): e1004110. https://doi.org/10.1371/journal.pgen.1004110

Michael A. J (1887) für Praktische Chemie, 2nd series, 35, 349–359.

Redman, K.L. (2006). Assembly of protein-RNA complexes using natural RNA and mutant forms of an RNA cytosine methyltransferase. Biomacromolecules 7, 3321–3326.

Safra M, Nir R, Farouq D, Vainberg Slutskin I, Schwartz S. 2017. TRUB1 is the predominant pseudouridine synthase acting on mammalian mRNA via a predictable and conserved code. Genome Res 27: 393–406.

Santi DV, Garrett CE, Barr PJ. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. Cell. Volume 33, Issue 1, May 1983, Pages 9-10

Santi, D. V., & Hardy, L. W. (1987) Biochemistry 26, 8599-8606.

Schaefer, M. *et al.* RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24**, 1590–1595 (2010).

Schoemaker, H. J. P. & Schimmel, P. R. Biochemistry 16, 5454–5460 (1977).

Schoemaker, H. J. P. & Schimmel, P. R. Biochemistry 16, 5461–5464 (1977).

Sharma S, Lafontaine DL (2015) 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. Trends Biochem Sci 40: 560–575

Spedaliere, C.J., and Mueller, E.G. (2004). Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine. RNA 10, 192–199.

Squires, J. E. et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. 40, 5023–5033 (2012).

Starzyk, R.M., S.W. Koontz, and P. Schimmel. 1982. A covalent adduct between the uracil ring and the active site of an aminoacyl tRNA synthetase. Nature 298: 136-140.

Stojkovic et al. (2018). miCLIP-MaPseq, a Substrate Identification Approach for Radical SAM RNA Methylating Enzymes. *J. Am. Chem. Soc.*, 2018, *140* (23), pp 7135–7143. DOI: 10.1021/jacs.8b02618

Van Haute L, Dietmann S, Kremer L, Hussain S, Pearce SF, Powell CA, Rorbach J, Lantaff R, Blanco S, Sauer S, Kotzaeridou U, Hoffmann GF, Memari Y, Kolb-Kokocinski A, Durbin R, Mayr JA, Frye M, Prokisch H, Minczuk M. Deficient methylation and formylation of mt-tRNA(Met) wobble cytosine in a patient carrying mutations in NSUN3. Nat Commun. 2016 Jun 30;7:12039.

Wu JC, Santi DV. Prog Clin Biol Res. 1985;198:119-29. On the mechanism and inhibition of DNA cytosine methyltransferases.

Yeh, Yun-Chi & Greenberg, G. R. J. biol. Chem. 242, 1307–1313 (1967).

Zaganelli S, Rebelo-Guiomar P, Maundrell K, Rozanska A, Pierredon S, Powell CA, Jourdain AA, Hulo N, Lightowlers RN, Chrzanowska-Lightowlers ZM, Minczuk M, Martinou JC (2017) The pseudouridine synthase RPUSD4 is an essential component of mitochondrial RNA granules. J Biol Chem 292:4519–4532

Zhao B.S. et al. (2017). Post-transcriptional gene regulation by mRNA modifications. Nat Rev Mol Cell Biol. 2017 Jan; 18(1): 31–42.